Automated Journalism: A Meta-Analysis of Readers' Perceptions of Human-Written in Comparison to Automated News

Graefe, Andreas; Bohlken, Nina

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Graefe, A., & Bohlken, N. (2020). Automated Journalism: A Meta-Analysis of Readers' Perceptions of Human-Written in Comparison to Automated News. *Media and Communication*, *8*(3), 50-59. https://doi.org/10.17645/mac.v8i3.3019

COGITATIO

Media and Communication (ISSN: 2183–2439)
2020, Volume 8, Issue 3, Pages 50–59
DOI: 10.17645/mac.v8i3.3019

Review

Automated Journalism: A Meta-Analysis of Readers' Perceptions of Human-Written in Comparison to Automated News

Andreas Graefe * and Nina Bohlken

Business Faculty, Macromedia University of Applied Sciences, 80335 Munich, Germany;
E-Mails: graefe.andreas@gmail.com (A.G.), ninabohlken@aol.com (N.B.)

* Corresponding author

Submitted: 14 March 2020 | Accepted: 21 May 2020 | Published: 10 July 2020

Abstract
This meta-analysis summarizes evidence on how readers perceive the credibility, quality, and readability of automated news in comparison to human-written news. Overall, the results, which are based on experimental and descriptive evidence from 12 studies with a total of 4,473 participants, showed no difference in readers' perceptions of credibility, a small advantage for human-written news in terms of quality, and a huge advantage for human-written news with respect to readability. Experimental comparisons further suggest that participants provided higher ratings for credibility, quality, and readability simply when they were told that they were reading a human-written article. These findings may lead news organizations to refrain from disclosing that a story was automatically generated, and thus underscore ethical challenges that arise from automated journalism.

Keywords
automated news; computational journalism; credibility; journalism; meta-analysis; perception; quality; review; robot journalism

Issue
This review is part of the issue "Algorithms and Journalism: Exploring (Re)Configurations" edited by Rodrigo Zamith (University of Massachusetts–Amherst, USA) and Mario Haim (University of Leipzig, Germany).

1. Introduction

Automated journalism, sometimes referred to as algorithmic journalism (Dörr, 2016) or robot journalism (Clerwall, 2014), alludes to the method by which algorithms are used to automatically generate news stories from structured, machine-readable data (Graefe, 2016).

The idea of news automation is not new. Half a century ago, Glahn (1970) described a process for automatically generating, what he called, "computer-produced worded weather forecasts." Basically, his idea was to create pre-written statements that describe different weather conditions, each of which corresponds to a particular output of a weather forecasting model (e.g., the combination of wind speed, precipitation, and temperature). This process is similar to today's template-based solutions offered by software providers in which a set of predefined rules are used to determine which prewritten statements are selected to create a story (Graefe, 2016).

Another domain that uses automated text generation is the financial news. In 2014, when the Associated Press gained much public attention for the decision to automate earnings reports (White, 2015), Thomson Financial (today part of Thomson Reuters) had already been automating such stories for nearly a decade (van Duyn, 2006).

It is no coincidence that weather and finance were the first applications to utilize news automation. In both domains, structured data, a requirement for news automation (Graefe, 2016), are available. Furthermore, data quality is high for these applications. Weather data are measured through sensors with relatively low measurement error. Likewise, the accuracy of company earnings or stock prices is critical for consumers of financial data.

What is new is the increasing abundance of structured and machine-readable data in many other domains. Governments are launching open data initiatives, sensors are constantly tracking environmental or health data, and users are leaving traces with virtually anything they do online. Such data can be used to generate automated news stories and thus serve as one of the technology's major drivers. Another important driver is economic pressure: News organizations need to save costs, increase news quantity (e.g., covering niche topics), and reach new target audiences (Graefe, 2016).

Promises of automation in increasing efficiency are manifold. As outlined by Graefe (2016), automating routine tasks has the potential to save resources and thus leave more time for journalists to do more important work, such as fact-checking or investigative reporting. Furthermore, automation can speed up news production and essentially enable publication as soon as the underlying data become available. Finally, algorithms tend to make fewer errors than human journalists and can personalize stories towards readers individual needs, and if necessary, in multiple languages.

Nevertheless, Dörr (2016) found news automation to be in an early market expansion phase at best. This situation does not seem to have changed much over the past four years. Providers of automated text generation still list few media organizations as their clients, although this may have to do with reasons of commercial confidentiality. That said, it is difficult to find regular text automation in high-profile publications, apart from the regularly cited one-off or experimental projects such as the Heliograf (*The Washington Post)* or ReporterMate (the Australian edition of *The Guardian*). Other major publications such as *The New York Times* stated that they are not planning to automatically generate news, despite having experimented with automation technology to personalize newsletters or moderate readers' comments (Peiser, 2019).

One reason for why news organizations refrain from using the technology, despite its economic potential, may be concerns that their readers would disapprove of automated news. According to the Modality–Agency–Interactivity–Navigability model (Sundar, 2008), readers may have a conflicting perception of automatically generated news. On the one hand, they may prefer human-written articles because they regard journalists as subject-matter experts (authority heuristic), or because they feel that they are communicating with a human rather than a machine (social presence heuristic). On the other hand, the machine heuristic suggests that readers regard automated news as free of ideological bias and thus more objective.

To answer such questions, researchers in different countries have conducted experimental studies to analyze how readers perceive automated news in comparison to human-written news. While sharing the common goal to better understand readers' perceptions of automated news, these studies often differed in their design. For example, some studies showed readers the same text and manipulated the byline as either written by a human or by an algorithm, whereas others revealed the true source of the articles. Yet another group of studies asked participants to rate either a human-written or an automated text, without revealing any information about who wrote the article.

The present meta-analysis summarizes available evidence on readers' perception of automated news to date, drawing on 11 articles, published in peer-reviewed journals between 2017 and 2020. Our goal is to give readers quick and easy access to prior knowledge. We provide an overview for which countries, domains, and topics evidence is available, which designs have been used to study the problem as well as how researchers recruited study participants. More importantly, we provide effect sizes aggregated across studies, while distinguishing between descriptive and experimental evidence as well as between effects that can be attributed to the article source (i.e., the author) and the message itself.

2. Method

2.1. Article Search

We included only studies published in scientific peer-reviewed journals in the English language. Studies had to provide experimental evidence on readers' perceptions of human-written news in comparison to automated news with respect to credibility, readability, and expertise. These are three of the four constructs that Sundar (1999) identified as central when people evaluate news content (the fourth one, representativeness, was omitted as it applies to news sections rather than single articles).

Our Google Scholar search for [('automated journalism' OR 'robot journalism') AND experiment AND perception] in October 2019 yielded 211 articles. After reading the title and abstract of each article, 34 articles were identified as potentially relevant and were thus read in full length by at least one of the authors. The articles by Jia (2020) and Tandoc, Yao, and Wu (2020) were added later. A total of 11 articles matched the inclusion criteria outlined above. Table 1 lists the 11 articles included in our meta-analysis. Three articles were published in *Digital Journalism*, and two articles each in *Journalism* and *Computers in Human Behavior.* The remaining four articles were published in four different journals, namely *International Journal of Communication*, *Journalism Practice*, *Journalism & Mass Communication Quarterly*, and *IEEE Access.*

2.2. Studies

We only included studies with a particular study design in our meta-analysis. These studies presented recipients with a short news story, in which either the author (journalist or algorithm), the attributed author (journalist or

algorithm), or both were experimentally manipulated. Recipients would then rate the article they had just read in terms of (at least one of the dimensions) credibility, quality, and readability.

Given that we were interested in readers' perceptions of human-written vs. automated news, we excluded experiments that used journalists as recipients (e.g., Jung, Song, Kim, Im, & Oh, 2017, Experiment 2) or analyzed hybrids of human-written and automated news (e.g., Waddell, 2019a). We also excluded studies that did not report effect sizes (e.g., Clerwall, 2014) or used a different experimental setup (e.g., Haim & Graefe, 2017, Experiment 2).

We ended up with 12 studies included in the 11 articles (cf. Table 1).

2.3. Coding

For each experimental study, one author coded study artifacts that related to the study participants, the stimulus material, the experimental design, and the study results (cf. Table 1). If the coder was uncertain regarding a particular coding, the issue was resolved by discussion with the second author. The coding sheet is available at the Harvard Dataverse (Graefe, 2020).

2.3.1. Participants

We coded the number of participants, age and gender distribution, the country/region participants came from as well as how participants were recruited. Across the 12 experiments, a total of 4,473 people participated, of which 50% were female. The average age was 36 years. Participants were from the USA (all of which were recruited through Amazon Mechanical Turk), Germany (recruited through the Sosci Panel administered by the German Communication Association), South Korea, China, Singapore, and other European countries.

2.3.2. Stimulus

We coded the domain of the news article, the article topic, as well as the article language. Sports news were most often used (eight studies), followed by financial (six) and political (four) news. Two studies focused on breaking news (earthquake alerts), and one study each used texts within the domains of entertainment and other news. Six experiments used articled written in English, two each in German and Korean, and one each in Finnish and Chinese.

2.3.3. Study Design

Table 1 shows the design for each study, particularly regarding our key variables of interest, namely who the actual author of the article was, and who was declared as the author (author attribution). In addition, Table 1 also lists additional experimental manipulations if available.

2.3.4. Outcome Variables

Across the 12 experiments, credibility was measured most often (nine times), followed by quality (eight times) and readability (five times). While the specific operationalization of the three constructs somewhat varied across studies, the measures used intend to capture the same basic constructs. Also, 8 of the 12 experiments reported effect sizes on a 5-point scale, three studies used a 7-point scale, and one study used a 10-point scale. For each outcome variable, we coded mean ratings and standard errors and/or standard deviations.

2.3.5. Effect Size Calculation

For each experimental comparison of human-written and automated news, we calculated Cohen's *d*, the standardized mean difference between the two groups, as:

$$d = \frac{\overline{M}_{HW} - \overline{M}_{A}}{SD_{pooled}}$$

where $\overline{M}_{HW}$ refers to participants' mean rating for perceptions of human-written news and $\overline{M}_{A}$ refers to mean ratings of automated news (Cohen, 1988). Hence, positive values for *d* imply that the human-written were rated better than automated news, and vice versa. Meta-analysis effect sizes were calculated as weighted (by the inverse of the variance) averages across the *d* values for the available studies. When referring to magnitudes of effects sizes, we adopted the descriptors suggested by Cohen (1988) and refined by Sawilowsky (2009), namely, zero effect ($d = 0$), very small effect ($0 < d < 0.2$), small effect ($0.2 \leq d < 0.5$), medium effect ($0.5 \leq d < 0.8$), large effect ($0.8 \leq d < 1.2$) very large effect ($1.2 \leq d < 2.0$), and huge effect ($d \geq 2.0$).

2.4. Types of Evidence

We distinguish between experimental and descriptive evidence in our analysis.

2.4.1. Experimental Evidence

Experimental evidence aims to establish causal effects by isolating the effects of the independent variable (i.e., the author or the attribution) through experimental manipulation.

Studies that aim to isolate the effect of the article source would show all recipients the same text (either written by a human or an algorithm). However, for some recipients, the text would be declared as written by a human, whereas for other recipients, that very same text would be declared as automatically generated.

Studies that aim to analyze the effect of the content (i.e., the message) would present recipients with either a human-written or an automated text but would not reveal the source (i.e., the texts had no byline).

Table 1. Experiments included in the meta-analysis.

			Participants					Stimulus								Experimental design (J = journalist, A = algorithm, U = unknown)									Outcome variables			
#	Citation	Exp.	N	% female	Avg. age	Country	Recruited	Language	Sports	Politics	Finance	Entertainment	Breaking	Other	Topic(s)	Author Attribution	U J	U A	J J	J A	A J	A A	J U	A U	Credibility	Quality	Readability	N-point scale
1	Wu (2019)		370	50	NA	USA	Commercial online panel (Amazon Mechanical Turk)	English	X	X	X				Not specified	2 (attribution) × 2 (author) × 3 (topic)			X			X	X	X	X			10
2	Jia (2020)	2	308	67	24	China	Social media snowball sampling (Wechat, Weibo, and Zhihu)	Chinese	X		X			X	Soccer, Basketball, Travel, Company reports, Conferences	2 (author) × 4 (topic)							X	X	X	X	X	5
3	Haim and Graefe (2017)	1	313	61	36	Germany	Non-commercial online panel (SoSci Panel)	German	X		X	X			Soccer, Stocks, Celebrities	2 (author) × 3 (topic)			X			X			X	X	X	5
4	Zheng, Zhong, and Yang (2018)		246	53	40	USA (154) & China (91)	Commercial online panel (Amazon Mechanical Turk)	English	X		X		X		Basketball, Stocks, Earthquake	2 (attribution) × 2 (media outlet) × 2 (culture) × 3 (topic)	X	X								X	X	7
5	Graefe, Haim, Haarmann, and Brosius (2018)		986	53	38	Germany	Non-commercial online panel (SoSci Panel)	German	X		X				Soccer, Stocks	2 (author) × 2 (attribution) × 2 (topic)			X	X	X	X			X	X	X	5
6	Wölker and Powell (2018)		300	60	28	Europe	Social media snowball sampling (Facebook, Twitter, and LinkedIn)	English	X		X				Basketball, Business (Forbes)	single factor (author) with 4 groups			X			X			X			5

Table 1. (Cont.) Experiments included in the meta-analysis.

			Participants					Stimulus							Experimental design (J = journalist, A = algorithm, U = unknown)									Outcome variables			
#	Citation	Exp.	N	% female	Avg. age	Country	Recruited	Language	Sports	Politics	Finance	Entertainment	Breaking	Other	Topic(s)	Author Attribution	U J	U A	J J	J A	A J	A A	J U	A U	Credibility	Quality	Readability	N-point scale
7a	Jung et al. (2017)	1	400	50	39	South Korea	Commercial online panel (Hankuk Research)	Korean (?)	X						Baseball	2 (author) × 2 (attribution)			X	X	X	X				X		5
7b		Pre-test	201	49	40	South Korea	Commercial online panel (Hankuk Research)	Korean (?)	X						Baseball	single factor (author) with two groups							X	X		X		5
8	Waddell (2018)		129	51	40	USA	Commercial online panel (Amazon Mechanical Turk)	English		X					Election polling	single factor (attribution) with 2 groups	X	X							X	X		7
9	Waddell (2019b)	1	612	47	38	USA	Commercial online panel (Amazon Mechanical Turk)	English		X					Khan Conflict, Paris Accord	3 (attribution) × 2 (media outlet) × 2 (topic)	X	X							X			7
10	Melin et al. (2018)		152	NA	NA	Finland	Commercial online panel	Finnish		X					Election results	single factor (author) with 4 groups							X	X	X	X	X	5
11	Tandoc et al. (2020)		420	41	38	Singapore	Commercial online panel	English					X		Earthquake	3 (attribution) × 2 (objectivity)	X	X							X			5
	Total		4,473	50	32	36			8	4	6	1	2	1			4	4	5	2	2	5	4	4	9	8	5	

2.4.2. Descriptive Evidence

Comparisons that provided descriptive evidence showed recipients news stories that were either written by a human or automatically generated, and truthfully declared the source. That is, human-written news would correctly be declared as written by a journalist, and automated news would correctly be labelled as generated by an algorithm. Then, the researchers would ask participants to rate these texts.

These comparisons do not allow for drawing causal inferences on the effects of the source or the message. However, for perceptions of credibility, researchers often use different scales (i.e., source credibility and message credibility), which were specifically designed to separate the effects. In contrast, no scales are available to distinguish the effects of the message and the source on perceived quality or readability. We were thus unable to separate the effects of source and message in these cases.

3. Findings

3.1. Main Effects

Figure 1 shows the main effects for each of the three constructs across all available comparisons, not differentiating between effects of the source and the message. Overall, there was no difference in how readers perceived the credibility of human-written and automated news (d = 0.0; SE = 0.02). Although human-written news were rated somewhat better than automated news with respect to quality, differences were small (d = 0.5; SE = 0.03). In terms of readability, the results showed a huge effect in that readers clearly preferred human-written over automated news (d = 2.8; SE = 0.04).

Interestingly, however, the direction of the effects for credibility and quality differed depending on the type of evidence. For both credibility (d = 0.3; SE = 0.03) and quality (d = 0.8; SE = 0.03), experimental evidence favored human-written over automated news. In comparison, descriptive studies showed the opposite effect: Automated news were favored over human-written news for both credibility (d = −0.5; SE = 0.04) and quality (d = −0.6; SE = 0.06).

3.2. Credibility

Figure 2 distinguishes between comparisons that provide evidence on the effects of the source and the effects of the message.

3.2.1. Source Credibility

One factor that may affect readers' perception of news is the source or, more specifically, the author. Overall, the results show a small difference between readers' perceptions of source credibility: human-written news were rated somewhat higher than automated news (d = 0.3; SE = 0.04). That said, the direction of effects differed again depending on the type of evidence. While experimental evidence showed a medium-sized effect in favor of human-written news (d = 0.5; SE = 0.04), descriptive evidence revealed a small effect in favor of automated news (d = −0.3; SE = 0.08).

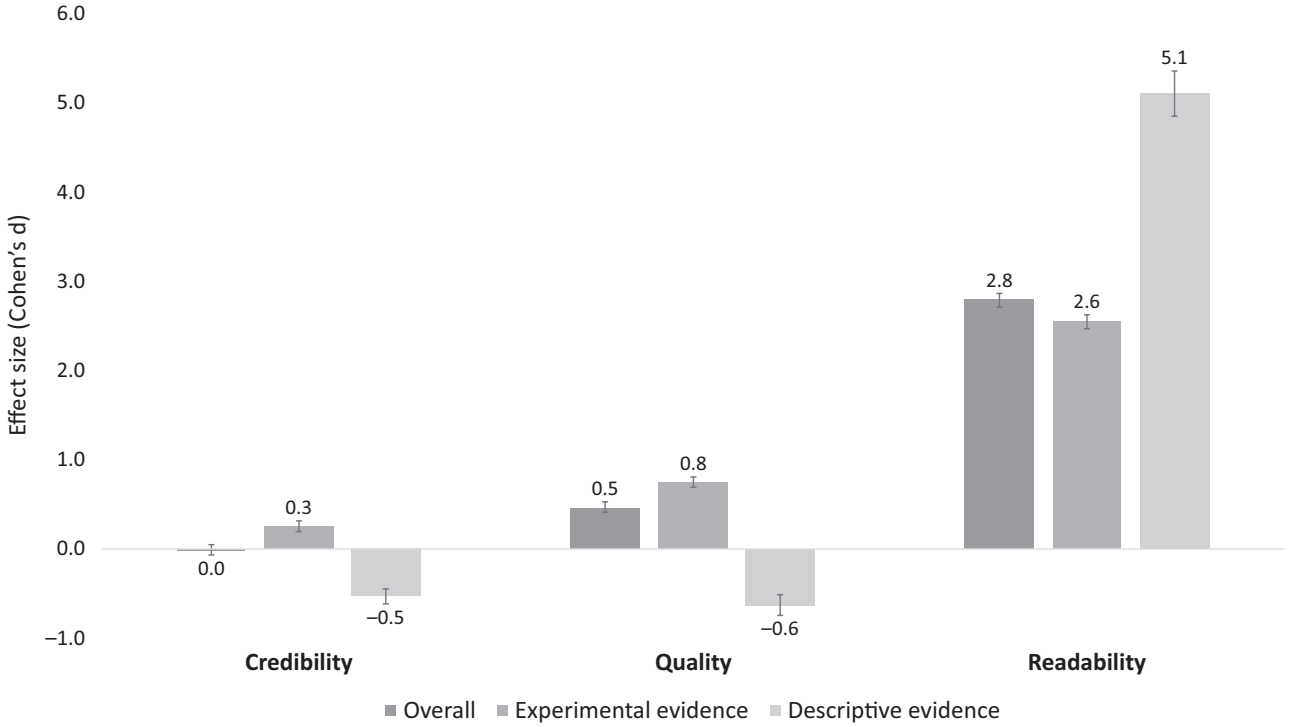

Figure 1. Main effects (standardized mean difference) for credibility, quality, and readability; by type of evidence. Note: Error bars show 95% confidence intervals.

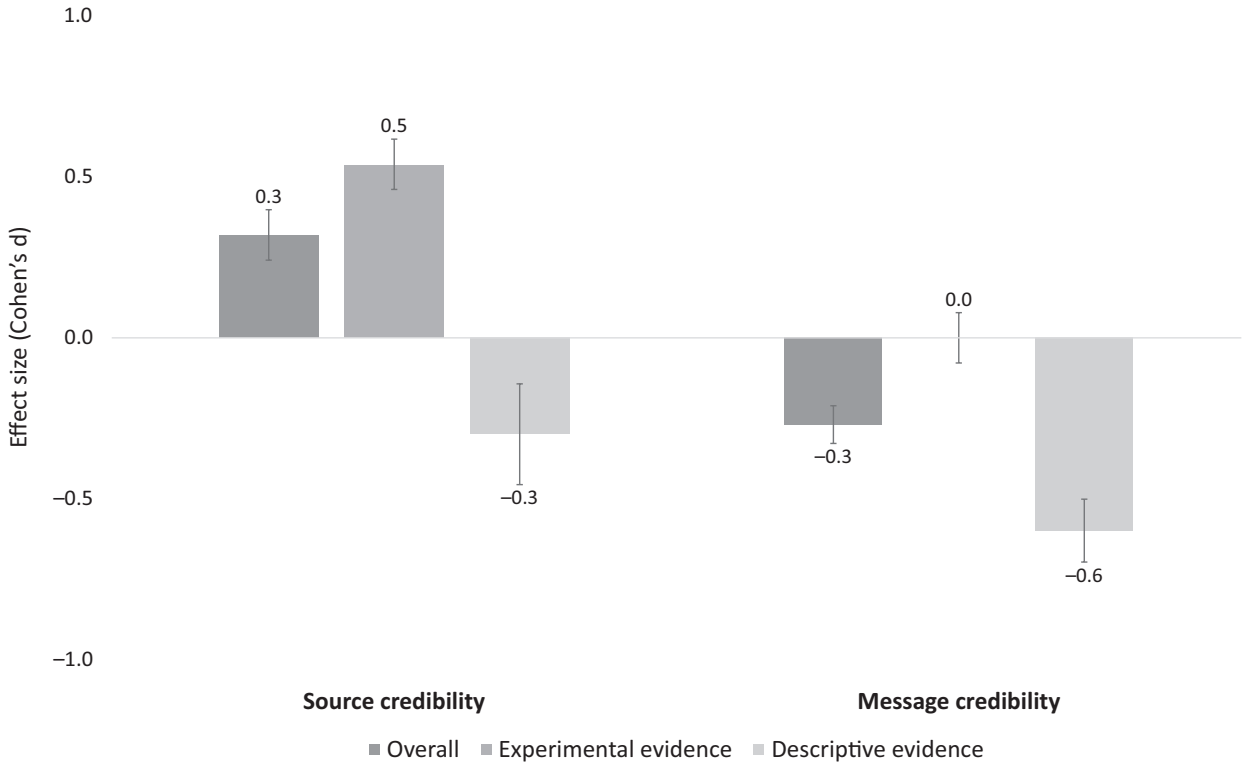

Figure 2. Standardized mean difference for credibility by type of evidence and effect. Note: Error bars show 95% confidence intervals.

3.2.2. Message Credibility

With respect to message credibility, automated news were rated somewhat more favorably across all comparisons ($d = -0.3$; SE = 0.03). Yet, again, the effect was solely carried by descriptive evidence ($d = -0.6$; SE = 0.05). Experimental evidence found no difference ($d = 0.0$; SE = 0.04).

3.3. Quality

Figure 3 distinguishes between experimental comparisons that provide evidence on the effects of the source and the effects of the message as well as descriptive evidence that does not allow for differentiating between effects of source and message on recipients' perceptions of quality.

Experimental evidence suggests that the article source has a small effect on perceptions of quality in that human-written news are rated somewhat better than automated news ($d = 0.3$; SE = 0.04). Experimental comparisons that provided evidence on the effects of the message found a very large effect in favor of human-written news ($d = 1.6$; SE = 0.05). Descriptive evidence, which does not allow for distinguishing between effects of the source and the message, found a medium-sized advantage for automated news with respect to perceived quality ($d = -0.6$; SE = 0.06).

3.4. Readability

Figure 4 distinguishes between experimental comparisons that provide evidence on the effects of the source and the message as well as descriptive evidence that does not allow for differentiating between effects of source and message on recipients' perceptions of readability.

Regardless of the type of evidence, the results showed a clear advantage for human-written articles. For experimental evidence on the effects of the source ($d = 1.8$; SE = 0.05) and the message ($d = 3.8$; SE = 0.07), effect sizes were very large and huge, respectively. Descriptive evidence on the combined effects of source and message showed a huge effect ($d = 5.1$; SE = 0.13).

4. Discussion

This meta-analysis aggregated available empirical evidence on readers' perception of the credibility, quality, and readability of automated news. Overall, the results showed zero difference in perceived credibility of human-written and automated news, a small advantage for human-written news with respect to perceived quality, and a huge advantage for human-written news with respect to readability.

One finding that stood out was the fact that the direction of effects differed depending on the type of evidence. Experimental evidence on the effects of the source found advantages for human-written news with

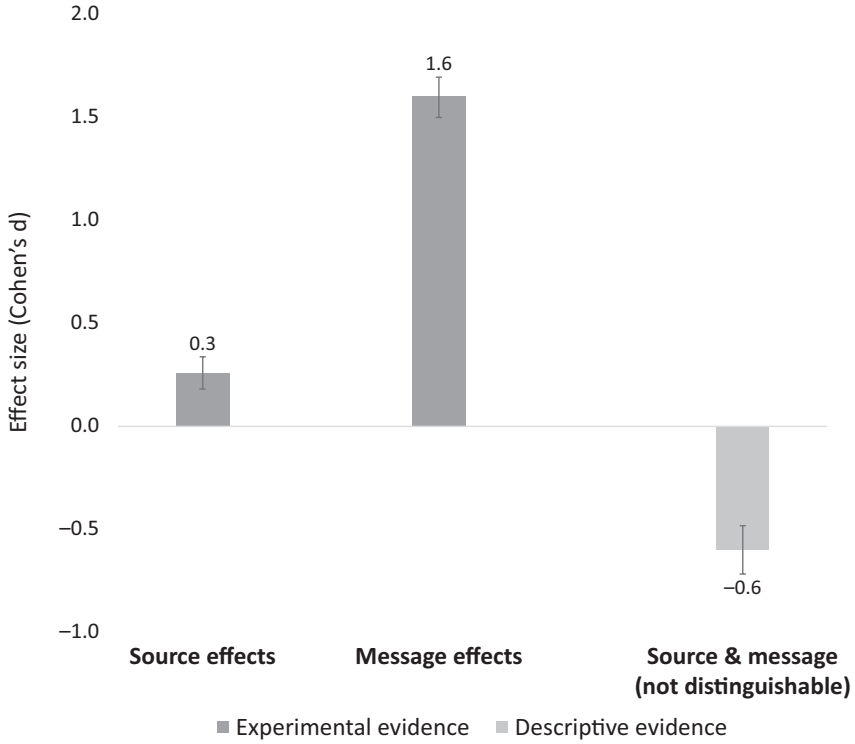

Figure 3. Standardized mean difference for quality by type of evidence and effect. Note: Error bars show 95% confidence intervals.

respect to quality (small effect), credibility (medium-sized), and readability (very large). In other words, regardless of the actual source, participants assigned higher ratings simply if they *thought* that they read a human-written article. The results thus support the authority heuristic and the social presence heuristic, while contradicting the machine heuristic (Sundar, 2008). Given these findings, news organizations may worry that their readers would disapprove of automated news and therefore refrain from disclosing that a story was auto-

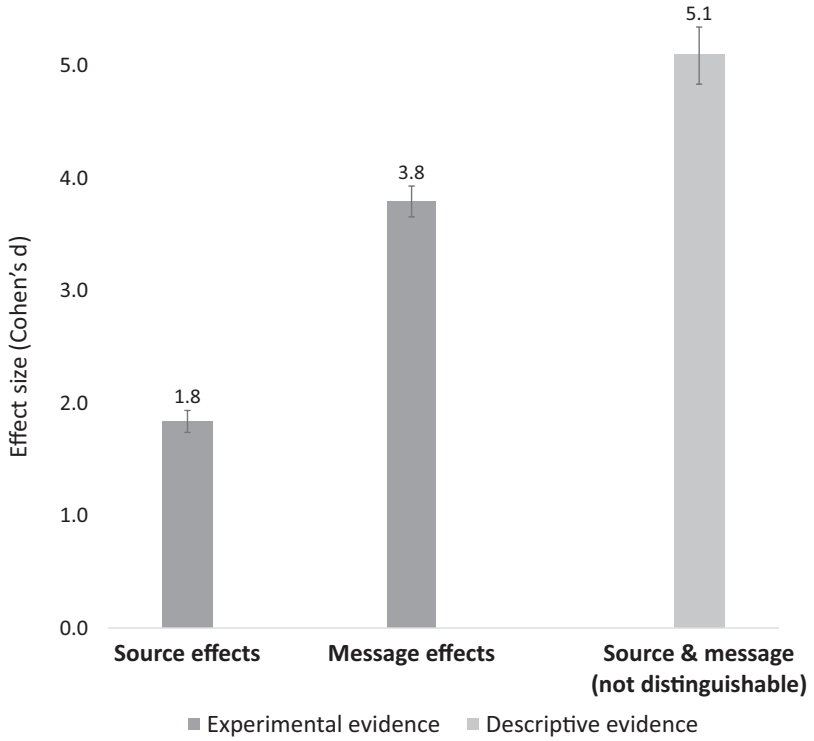

Figure 4. Standardized mean difference for readability by type of evidence and effect. Note: Error bars show 95% confidence intervals.

matically generated (e.g., by not providing a byline). This underscores the ethical challenges that arise from automated journalism (Dörr & Hollnbuchner, 2017).

Experimental evidence further showed advantages for human-written news with respect to the effect of the message (i.e., the article content). If participants did not know what they were reading, they assigned higher ratings to human-written news compared to automated news with respect to quality (very large effect) and readability (huge effect). There was, however, no effect for credibility. Obviously, these results depend entirely on the actual articles used in these comparisons. We thus refrain from deriving any conclusions or practical implications from these results and expect that the human written articles in these particular comparisons may have simply been better than the automated counterparts. The extent to which these articles are representative of the relative quality of automated and human-written news at the time is unclear.

In contrast, descriptive evidence showed opposite results with respect to how article source and message affect perceptions of credibility and quality. That is, automated news were perceived as more credible and of higher quality than the human-written counterparts in studies that asked recipients to evaluate articles whose source was truthfully declared. Given that these studies do not allow for making causal inferences, it is difficult to draw practical implications. In particular, any differences in effect sizes could simply be due to differences in the actual quality of the articles themselves.

Our analysis thus demonstrates the importance of distinguishing between the type of evidence (descriptive vs. experimental) as well as the origin of the effect (source and message). Otherwise, interesting findings, such as the positive effect of human authors on people's perceptions may get lost in the aggregate. That said, effects of the source with respect to both perceived credibility and quality were small. News organizations may not need to worry too much that readers could perceive automated news as less credible, or more generally as being of lower quality, than human-written news—assuming of course that the articles' actual quality is similar.

Differences with respect to readability, however, were huge. On the one hand, one could assume that poor readability is a critical barrier for readers' willingness to consume automated news. On the other hand, it should be noted that automation is most useful for routine and repetitive tasks, for which one needs to write a large number of stories (e.g., weather reports, corporate earnings stories). Such routine writing is often little more than a simple recitation of facts that neither requires flowery narration nor storytelling. In fact, in certain domains such as financial news, sophisticated writing may even be harmful, as consumers generally want the hard facts as quickly and clearly as possible. Another potential benefit of automation is the possibility to cover topics for very small target groups, for which previously no news were available (e.g., lower league games for niche sports, earthquake alerts, fine dust monitoring, etc.). For such topics, readers may be happy if they get any news at all. As a result, they may not be too concerned with readability, especially with how the construct is commonly measured (e.g., with items such as 'entertaining,' 'interesting,' 'vivid,' or 'well written') in the literature. Future research should analyze perceptions of readers that represent the actual target group (i.e., people who would actually consume automated news).

Needless to say, our results provide merely a snapshot of the current state of news automation, drawing on evidence from 11 articles published between 2017 and 2020. Readers' perceptions may change over time, and they may change fast. Assuming that automated news becomes more common, readers would get more accustomed to such content, which could ultimately affect their perceptions. Also, the technology for creating automated news, as well as the availability of data, is likely to further improve over time, which we expect to positively affect both the quality and readability of automated news. Advances in statistical analysis, in combination with more data, should make it possible to add more context (e.g., adding weather data to exit polling texts) and analytical depth (e.g., by analyzing historical data, making predictions, etc.), which should improve the perceived quality of such texts. Similarly, we would expect natural language generation to further improve, with positive effects on perceived readability. Future research should continue monitoring readers' perception of automated news, especially if and how improvements in the objective quality of the texts affect their perceived quality.

The latter relationship has generally been overlooked in research to date. Available studies have merely analyzed if, and to what extent, readers' perceptions of automated and human-written news differ. Yet, we do not know which factors drive these perceptions. What is it that makes an article perceived as more or less credible or readable? Such information would be valuable for developers of automated news to improve the (perceived) quality of the texts.

Acknowledgments

Jamie Graefe edited the article.

Conflict of Interests

The authors declare no conflict of interests.

References

Clerwall, C. (2014). Enter the robot journalist. *Journalism Practice*, *8*(5), 519–531.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dörr, K. N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, *4*(6), 700–722.

Dörr, K. N., & Hollnbuchner, K. (2017). Ethical challenges of algorithmic journalism. *Digital Journalism, 5*(4), 404–419.

Glahn, H. R. (1970). Computer-produced worded forecasts. *Bulletin of the American Meteorological Society*, *51*(12), 1126–1132.

Graefe, A. (2016). *Guide to automated journalism*. New York, NY: Tow Center for Digital Journalism.

Graefe, A. (2020). Data for automated journalism: A meta-analysis of readers' perceptions of human-written vs. automated news. *Harvard Dataverse*. doi:10.7910/DVN/Q4LLYW

Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, *19*(5), 595–610.

Haim, M., & Graefe, A. (2017). Automated news: Better than expected? *Digital Journalism*, *5*(8), 1044–1059.

Jia, C. (2020). Chinese automated journalism: A comparison between expectations and perceived quality. *International Journal of Communication*, *14*, 2611–2632.

Jung, J., Song, H., Kim, Y., Im, H., & Oh, S. (2017). Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior*, *71*, 291–298.

Melin, M., Bäck, A., Södergård, C., Munezero, M. D., LeppäNen, L. J., & Toivonen, H. (2018). No landslide for the human journalist: An empirical study of computer-generated election news in Finland. *IEEE Access*, *6*, 43356–43367.

Peiser, J. (2019). The rise of the robot reporter. *The New York Times*. Retrieved from https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–599.

Sundar, S. S. (1999). Exploring receivers' criteria for perception of print and online news. *Journalism & Mass Communication Quarterly*, *76*(2), 373–386.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 70–100). Cambridge, MA: MIT Press.

Tandoc, E. C., Jr., Yao, L. J., & Wu, S. (2020). Man vs. machine? The impact of algorithm authorship on news credibility. *Digital Journalism*, *8*(4), 548–562.

van Duyn, A. (2006). Computers write news at Thomson. *Financial Times*. Retrieved from https://www.ft.com/content/bb3ac0f6-2e15-11db-93ad-0000779e2340

Waddell, T. F. (2018). A robot wrote this? *Digital Journalism*, *6*(2), 236–255.

Waddell, T. F. (2019a). Attribution practices for the man-machine marriage: How perceived human intervention, automation metaphors, and byline location affect the perceived bias and credibility of purportedly automated content. *Journalism Practice*, *13*(10), 1255–1272.

Waddell, T. F. (2019b). Can an algorithm reduce the perceived bias of news? Testing the effect of machine attribution on news readers' evaluations of bias, anthropomorphism, and credibility. *Journalism & Mass Communication Quarterly*, *96*(1), 82–100.

White, E. M. (2015). Automated earnings stories multiply. *Associated Press*. Retrieved from https://blog.ap.org/announcements/automated-earnings-stories-multiply

Wölker, A., & Powell, T. E. (2018). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*. https://doi.org/10.1177/1464884918757072

Wu, Y. (2019). Is automated journalistic writing less biased? An experimental test of auto-written and human-written news stories. *Journalism Practice*. https://doi.org/10.1080/17512786.2019.1682940

Zheng, Y., Zhong, B., & Yang, F. (2018). When algorithms meet journalism: The user perception to automated news in a cross-cultural context. *Computers in Human Behavior*, *86*, 266–275.

About the Authors

Andreas Graefe is Professor of Management at Macromedia University of Applied Sciences in Munich, Germany. He has been studying automated journalism since 2014 and published several articles on that topic, including the *Guide to Automated Journalism* published by the Tow Center for Digital Journalism at Columbia University. Andreas is also an expert in forecasting, particularly election forecasting.

Nina Bohlken is currently finishing up her BA in journalism at Macromedia University of Applied Sciences in Munich.